

Lecture 5

Cryptography Basics

CS3690 Network Security
 Summer Quarter, 2000
 C. Irvine

Objectives

- Concepts from Classical Cryptography
- Feistel Ciphers
- Block Ciphers

Summer Quarter, 2000 C. Irvine; NPS CISR 2

Definitions

- **Encryption** - encode.
- **Decryption** - decode.
- **Plaintext** - information to be sent
- **Ciphertext** - encrypted information
- **Cryptology** - study of encryption and decryption.
- **Cryptography** - using encryption to conceal text.
- **Cryptanalysis** - the breaking of secret writing.

Summer Quarter, 2000 C. Irvine; NPS CISR 3

Cryptosystem

- A cryptosystem is a five-tuple (P, C, K, E, D) , where the following conditions are satisfied
 1. P is a finite set of possible plaintexts
 2. C is a finite set of possible ciphertexts
 3. K , the keyspace, is a finite set of possible keys
 4. For each k in K , there is an encryption rule e_k in E and a corresponding decryption rule d_k in D . Each $e_k : P \rightarrow C$ and $d_k : C \rightarrow P$ are functions such that $d_k(e_k(x)) = x$ for every plaintext x in P .
- **Observations**
 1. if we have a string $x = x_1 x_2 x_3 \dots x_n$
 2. when Alice encrypts it there will be a resulting string $y = y_1 y_2 y_3 \dots y_n$
 3. We need the encryption function to be one-to-one so it is never the case that $y = e_k(b) = e_k(b')$ where $x_i \neq x_j$
 4. If $P = C$, then the elements of P are rearranged (permuted) by the encryption function.

Summer Quarter, 2000 C. Irvine; NPS CISR 4

Classical Cryptographic Systems

- Use substitution and transposition
- Have a specific number of keys
- Handle plaintext either in blocks or streams
- Encryption schemes fall into two major categories
 - ★ Unconditionally secure
 - ★ Computationally secure
 - Cost to break cipher exceeds value
 - Time to break cipher exceeds useful lifetime
- **Types of Attacks**
 - ★ Ciphertextonly
 - ★ Known plaintext (and sometimes probable plaintext)
 - ★ Chosen plaintext
 - ★ Chosen ciphertext - uncommon
 - ★ Chosen text - uncommon

Summer Quarter, 2000 C. Irvine; NPS CISR 5

Elements of a Good Cryptosystem

- Each encryption function, x , and decryption function, y , should be computationally efficient
- If an opponent has a copy of the cipher text, it should be impossible to determine either the key used or the plain text string

Summer Quarter, 2000 C. Irvine; NPS CISR 6

Cryptanalysis

- Oscar's chore as a cryptanalyst:
 - ★ Break a single message.
 - ★ Recognize patterns in order to develop decryption algorithm.
 - ★ Find general weakness in encryption algorithm.
- Breakable encryption algorithm
 - ★ Feasible given sufficient time and data.
 - ★ Brute force attack is usually impractical.
 - ★ Estimates of breakability - based on current technology.
 - ★ Although scheme is based on a 'hard' problem doesn't mean that the cryptanalyst will attempt to solve it that way.

Summer Quarter, 2000

C. Irvine; NPS CISR

7

Cryptanalyst's Tools

- Traditional Tools
 - ★ Letter frequency data
 - ★ Prefix/suffix lists
 - ex ant-, inter-, post-, -ale, -ing
 - ★ Letter pair/triple lists
 - ex -re-, -th-, -en-, -de-, -ion-, -ive-, -ble-
 - ★ Common pattern lists
 - ex -eek-, -oot-, -our-

Summer Quarter, 2000

C. Irvine; NPS CISR

8

Number Theory

- Modulus
- Take a and b to be integers and m to be a positive integer.
- when we write (mod m)
- if m divides b-a.
- When we describe (mod m) in words we say that "a is congruent to b modulo m." The modulus is the integer m.
- We can define arithmetic modulo m:
 - Z_m is defined to be the set $\{0, 1, \dots, m-1\}$
- with two operations +, addition, and x, multiplication, which work the same as in the usual sense except that they are reduced modulo m.

Summer Quarter, 2000

C. Irvine; NPS CISR

9

Rules for Modular Arithmetic

- Arithmetic for addition and multiplication modulo m
 1. addition is closed: for any a, b in Z_m , $(a+b)$ is in Z_m
 2. addition is commutative: for any a, b in Z_m , $a+b = b+a$
 3. addition is associative: for any a, b, c in Z_m , $a + (b+c) = (a+b) + c$
 4. 0 is the additive identity: for any a in Z_m , $a + 0 = 0 + a = a$
 5. the additive inverse of any a in Z_m is m-a, for example $a + (m-a) = (m-a) + a = 0$ for any a in Z_m
 6. multiplication is closed: for any a, b in Z_m , ab in Z_m ,
 7. multiplication is commutative: for any a, b in Z_m , $ab = ba$
 8. multiplication is associative: for any a, b, c in Z_m , $(ab)c = a(bc)$
 9. 1 is the multiplicative identity: for any a in Z_m , $a \times 1 = 1 \times a = a$
 10. multiplication distributes over addition: for any a, b, c in Z_m , $(a+b)c = (ac) + (bc)$ and $a(b+c) = (ab) + (ac)$

Summer Quarter, 2000

C. Irvine; NPS CISR

10

More Number Theory

1. An algebraic structure having properties 1, 3, 4, and 5 is called a group with respect to addition. Property 4 makes the structure an abelian group
2. Properties 1-10 go further and we can call Z_m a ring.
3. Examples of rings
 - ★ integers
 - ★ real numbers
 - ★ complex numbers
- For cryptographic algorithms using modular arithmetic we are interested in a finite rings
- So applying our rules modulo 3, we have
 - $0 \bmod 3 = 0$; $1 \bmod 3 = 1$; $2 \bmod 3 = 2$;
 - $3 \bmod 3 = 0$; $4 \bmod 3 = 1$; $5 \bmod 3 = 2$;
 - $6 \bmod 3 = 0$; $7 \bmod 3 = 1$; ...

Summer Quarter, 2000

C. Irvine; NPS CISR

11

Shift Cipher - Simple Substitution Cipher

- Definition
 - Let $P = C = K = Z_{26}$ for 0..le. K..le. 25
 - $e_k(x) = x + K \bmod 26$
 - and
 - $d_k(x) = x - K \bmod 26$
 - where x,y are in Z_{26}
- Representation of characters
 - ★ Algorithms are usually mathematical in nature.
 - ★ Using the representation below, arithmetic operations can be performed on the letters:

letter:	A	B	C	D	E	F	G	H	I	J	K	L	M
code:	0	1	2	3	4	5	6	7	8	9	10	11	12
letter:	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
code:	13	14	15	16	17	18	19	20	21	22	23	24	25

Summer Quarter, 2000

C. Irvine; NPS CISR

12

Example

1. let $K = 11$ with plain text
 we will meet at midnight
 22 4 22 8 11 12 4 19 0 19 12 8 3 13 8 6 7 19
2. first we convert everything to numbers
3. add 11 and take the modulus 26 of each sum
 7 15 7 19 22 22 23 15 15 4 11 4 23 19 24 17 18 4
4. then we convert this to alphabetic characters
 hphtwxppelextoytrse

Summer Quarter, 2000

C. Irvine; NPS CISR

13

Shift Ciphers - Caesar Cipher

- Each letter translated a fixed number of letters.
 - Caesar used a shift of three (i.e. $c_i = E(p_i) = p_i + 3$)
- Advantage
 - Easy to use in the field.
- Disadvantage
 - Pattern simple to determine.
- Cryptanalysis of Caesar Cipher
 - Obvious break between words. Of course, this is not typical of encryption schemes. Usually the spaces are omitted.
 - Double letter pairs easy to spot.
 - Repeated letters.

Summer Quarter, 2000

C. Irvine; NPS CISR

14

Shift Cipher - Playfair Cipher

- Keyword starts 5x5 matrix

m	o	n	a	r
c	h	y	b	d
e	g	u	j	k
i	p	q	s	t
x	w	v	z	
- Work on digrams of plaintext (total possible is $26 \times 26 = 676$)
 - repeated letters in plaintext have x inserted between them. Thus good becomes goxod
 - plaintext letters in the same row are replaced by the letter to the right: ar becomes rm
 - plaintext letters in the same column are replaced by the letter below: mu is encrypted as cm
 - all others are replaced by row/column combinations: hs becomes bp and ea becomes jm

Summer Quarter, 2000

C. Irvine; NPS CISR

15

Monoalphabetic Ciphers

- Advantages of monoalphabetic ciphers
 - Can be performed by direct lookup.
 - Time to encrypt message of n characters is proportional to n .
- Frequency Distributions
 - Recall the simple message that we deciphered, which had been encrypted with a simple Caesar cipher. Although we took it for granted at the time, we had used some knowledge of our own language to assist us in deciphering the text. Specifically, letter pairs and small words to help us guess a solution.
 - We'll use the following example to see how letter frequencies can betray our secrets.

Summer Quarter, 2000

C. Irvine; NPS CISR

16

Cryptanalysis of Monoalphabetic Ciphers

- Frequency Distribution
 - Note that in the cipher text (out of 166 letters):

vowel	A	E	I	O	U
count	0	2	2	2	4
percent	0.0	1.2	1.2	2.41	4.79
 - Note that in the plain text (out of 166 letters):

vowel	A	E	I	O	U
count	11	26	16	9	5
percent	6.59	15.56	9.58	5.39	2.99
 - Note that in the in English language:

vowel	A	E	I	O	U
percent	7.49	14.0	6.65	7.37	3.0

Summer Quarter, 2000

C. Irvine; NPS CISR

17

Are Monoalphabetic Ciphers Secure?

- At First Glance:
 - 26! possible encipherments.
 - At one decipherment per microsecond it would take 1000 years to test all 26! decipherments by brute force.
- The Answer - NO!
- In a long message letter frequencies betray the text
 - Observation #1
 - An encryption based on a hard problem is not secure just because of the difficulty of the problem.
 - Observation #2
 - An encryption algorithm must be regular in order to be algorithmic and therein lies its weakness.
 - Observation #3
 - A security measure must be strong enough to keep out the attacker for the life of the data only.

Summer Quarter, 2000

C. Irvine; NPS CISR

18

Polyalphabetic Substitution Ciphers

■ Example

1. Create two alphabets by using Multiplicative Modulus Permutations.
2. Alternate alphabets while encrypting or decrypting the text.
odd positions use: $f(i) = (3 * i) \bmod 26$
even positions use: $f(i) = ((5 * i) + 13) \bmod 26$

Alphabet for odd positions:

ABCDEF GHIJ KLMNOP QRSTUV WXYZ
ADGJMPSVYBEHKNQ TWZCFILORUX

Alphabet for even positions:

ABCDEF GHIJ KLMNOP QRSTUV WXYZ
NSXCHMRWBGLOVAFKPUEJOTYDI

■ Advantages of Polyalphabetic Substitutions:

- ★ Flattens letter frequencies.
- ★ Double letter pairs not so obvious.
- ★ Prefix tables become more complicated.

Summer Quarter, 2000

C. Irvine; NPS CISR

19

Cryptanalysis of Polyalphabetic Ciphers

■ Vigenere Tableaux - highly sophisticated polyalphabetic cipher

■ Cryptanalysis of Polyalphabetic Substitutions

- ★ Appears to be more secure! Although it is more complex, polyalphabetic ciphers are not immune from breaking.
- ★ They can be broken using two tools.
 - Kasiski method
 - Index of coincidence

Summer Quarter, 2000

C. Irvine; NPS CISR

20

Kasiski Method for Repeated Patterns

■ Relies on the regularity of the English language.

- ★ Letters, letter groupings and words are repeated.

- Endings: -th, -ing, -ed, -ion, -tion
- Beginnings: Im-, in-, un-, re-
- Patterns: -eek-, -ool-, -our-
- Words: for, to, of, with, are, is

Observation #1

If a message is encoded with n alphabets in cyclic rotation, and if a particular word group appears k times in a plaintext message, it should be encoded approximately k/n times from the same alphabet.

Observation #2

If a plaintext phrase is enciphered the same way twice, the key must have gone through a whole number of rotations and be back at the same point.

Observation #3

The distance between the repeated patterns must be a multiple of the keyword length.

Summer Quarter, 2000

C. Irvine; NPS CISR

21

Steps to Kasiski Method

■ Steps to Kasiski Method

1. Identify repeated patterns of three or more characters.
2. Jot down the starting position for each instance of the pattern.
3. Compute the difference between the starting points.
4. Determine all factors of each difference.
5. Key length will be one of the factors appearing in step [4]

Repeated patterns over 3 characters not accidental
Likelihood of two 4-letter pattern by accident $1/26^4$

■ Example of Kasiski Method

dicke nsdic kensd icken sdick ensdi ckens dicke nsdic kensd
ITWAS THERE STOFT IMESI TWAST HEWOR STOFT IMESI TWAST HEAGE
icken sdick ensdi ckens dicke nsdic kensd icken sdick ensdi
OFWIS DOMIT WASTH EAGEO FFOOL ISHNE SSITW ASTHE EPOCH OFBEL
ckens dicke nsdic kensd
IEFIT WASTH EEPFC HOPIN

Observe - Phrase IT WAS THE is enciphered with the keyword nsdicken three times

Start	Distance	Factors
20	---	---
83	63 (83 - 20)	3, 7, 9, 21, 63
104	21 (104 - 83)	3, 7, 21

Note: length("dickens") = 7

Summer Quarter, 2000

C. Irvine; NPS CISR

22

Index of Coincidence

- Try to divide the message into pieces enciphered with the same alphabet (place letters in respective alphabet buckets).
- Continuing The Example
- Looking at the factored distances in our example, we first assume a key length of 3.
S1 = {c1, c4, c7, c10...}
S2 = {c2, c5, c8, c11...}
S3 = {c3, c6, c9, c12...}
- S1, S2 and S3 represent all those characters enciphered with the same alphabet and should have frequency distributions similar to English and distributions similar to the other sets.
 - ★ The index of coincidence is a measure of the variation between frequencies in a distribution.
 - ★ If the distribution of all the letters of the alphabet was absolutely flat they would be distributed with an equal probability of occurrence of $1/26$ or approximately 0.0384.
 - ★ If we look at a plot of how letters in the English language are actually distributed we can then look at the variance (a measure of roughness) of the whole distribution rather than the individual letters.
 - ★ If the amount of ciphertext is large and the plaintext letter distribution is typical of the language then the IC can be used to predict the number of alphabets used.

Summer Quarter, 2000

C. Irvine; NPS CISR

23

More Index of Coincidence

■ Number of enciphering Alphabets vs IC

1 2 3 4 5 10 large
.068 .052 .047 .044 .043 .042 .038

- Going back to the example we would examine all three of our sets S1, S2 and S3 to see if the ICs of all three were close to 0.068. If they were not then we would try our next guess of a key of length 7 (which in this case is correct).

A useful observation at this point is that the cryptanalysis involves computation, but unless the number of alphabets approaches a truly huge number, an adversary with lots of computing power is going to be able to crack the code

Summer Quarter, 2000

C. Irvine; NPS CISR

24

Perfect Substitution Cipher

- How to flatten distribution so IC of ciphertext is close to 0.038?
 - By using a very large number of alphabets.
 - Ideally, an infinite number in order to produce a perfectly flat distribution.
- One Time Pad
 - Large nonrepeating keys written on a pad.
 - Since each key is different, a one-time pad equivalent one large key.
- Example
 - If the keys are 20 characters long and a message was 300 characters long, the sender would have to use 15 keys ($20 \times 15 = 300$). Then the sender would write them one at a time above the plaintext and encipher with a Vigenere Table.
- Problems:
 - Key printing
 - Distribution
 - Storage

Summer Quarter, 2000

C. Irvine; NPS CISR

25

Perfect Substitution Cipher: Vernam Cipher

- Uses a long nonrepeating sequence of numbers combined with the plaintext.
- This system is immune from cryptanalytic attack since ciphertext does not give away the key.
 - Vernam used a long punched tape that fed into a teletype machine.
 - Each tape was used only once.
- Method
 - Use the enumeration set position for the character p_i (0..25).
 - Add a random number to it ($p_i + \text{random}$).
$$c_i = (p_i + \text{random}) \bmod 26.$$

Summer Quarter, 2000

C. Irvine; NPS CISR

26

Binary Vernam Cipher

- 1. Take binary stream (the plaintext).
 - 2. XOR random binary stream (the key).
 - 3. Produce binary ciphertext.
- Example
- | | | | | |
|----------|-------------|--------------------|------------------|--|
| Encrypt | plaintext | 1 0 1 1 0 1 | key | |
| | 1 0 1 1 1 1 | binary cipher text | 0 0 0 0 1 0 | |
| De crypt | cipher text | 0 0 0 0 1 0 | | |
| | key | 1 0 1 1 0 1 | binary plaintext | |
- Vernam Cipher Depends upon RGN

Summer Quarter, 2000

C. Irvine; NPS CISR

27

Transposition (Permutation) Ciphers

- Method
 - Plaintext characters are arranged into columns.
 - The columns are then permuted.
- Example
 - Plaintext message: THIS IS A MESSAGE TO SHOW HOW A COLUMNAR TRANSPOSITION WORKS
 - THIS IS A MESSAGE TO SHOW HOW A COLUMNAR TRANSPOSITION WORKS
 - Ciphertext message: TSSOH OANIW HAAO LRSTO IMGHW UTPIR SEEAW MROOK ISTWC NASNS
- Advantages / Disadvantages
 - Simple method but effective.
 - Work per character is constant but time dependent on message length.
 - Requires buffered storage the size of which depends on length of message.
 - Delay in transmission since the entire message must be in memory before the encryption process can take place.
 - Not appropriate for long messages.

Summer Quarter, 2000

C. Irvine; NPS CISR

28

Cryptanalysis of Permutation Ciphers

- Pattern Analysis
 - Letter frequency looks like monoalphabetic cipher.
 - Frequency tables that are useful.
 - Digrams - letter pairs (-re-, -th-, -ed-)
 - Trigrams - letter triples (-ene-, -ion-, -ite-)
- The Problem:
 - Where are the adjacent columns in the ciphertext?

Summer Quarter, 2000

C. Irvine; NPS CISR

29

Analysis for Digrams and Trigrams

- From our previous example:

TSSOH OANIW HAAO LRSTO IMGHW UTPIR

 - Break the cipher text into columns and record digrams

T	S	S	O	H	O	A	N	I	W	H	A	A	O	L	R	S	T	O	I	M	G	H	W	U	T	P	I	R
S	I	T	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
S	W	S	W	T	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W
H	A	O	A	S	A	S	A	S	A	S	A	S	A	S	A	S	A	S	A	S	A	S	A	S	A	S	A	S
O	A	H	A	O	A	S	A	S	A	S	A	S	A	S	A	S	A	S	A	S	A	S	A	S	A	S	A	S
A	S	O	S	H	S	O	S	H	S	O	S	H	S	O	S	H	S	O	S	H	S	O	S	H	S	O	S	H
L	A	O	O	R	O	L	A	O	O	R	O	L	A	O	O	R	O	L	A	O	O	R	O	L	A	O	O	R
L	A	O	O	R	O	L	A	O	O	R	O	L	A	O	O	R	O	L	A	O	O	R	O	L	A	O	O	R
R	R	R	R	A	R	R	R	A	R	R	R	A	R	R	R	A	R	R	R	A	R	R	R	A	R	R	R	A
S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S
T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O
- Functional Relationship
 - $c_i = E(p) = \text{ColLen} \cdot ((q-1) \bmod_{\text{No_Col}}) + ((q-1) / \text{No_Col} + 1)$

Summer Quarter, 2000

C. Irvine; NPS CISR

30

More on Transpositions

■ Double Transpositions

- ★ Use two columnar transpositions applied one after the other.
- ★ Use different number of columns for each transposition.
- ★ First transposition displaces adjacent letters.
- ★ Second transposition breaks up short series adjacencies from the first.

■ Cryptanalysis - Functional Relationship

- ★ A functional relationship still exists, although it is much more complex!

Summer Quarter, 2000

C. Irvine; NPS CISR

31

Old Fashioned Stream and Block Ciphers

■ Stream Ciphers (Substitution)

- ★ **Advantages**
 - Speed of transformation.
 - How error propagation (errors effect only single character).
- ★ **Disadvantages**
 - Information not diffused.
 - Susceptible to modification and insertion.

■ Block Ciphers (Transposition)

- ★ **Advantages**
 - Information is diffused (one block may depend on another).
 - Immune to insertions.
- ★ **Disadvantages**
 - Very Slow!
 - Error propagation (errors affect the entire block).

Summer Quarter, 2000

C. Irvine; NPS CISR

32

DES is Insecure

- 56-bit key is vulnerable to brute force search
- German Court Rule's DES inadequate in September 1998
<http://www.thestandard.net/article/display/0,1151,1780,00.html>
- Distributed cracking Parallelizeswork
<http://www.eff.org/descracker.html>
- 75 bit keys is minimal to protect data today
- To protect data for 20 years, use 90-bit keys
- Moore's Law
 - ★ machine speed doubles every 18 months
 - ★ In 10 years machine to break DES \$2000

Summer Quarter, 2000

C. Irvine; NPS CISR

33